

**METHOD AND APPARATUS FOR
MONITOR, SEARCH AND NOTIFICATION
IN A NETWORK**

FIELD OF THE INVENTION

This invention relates to a method and apparatus for monitoring and searching a Network, such as the Internet, and for notifying one or more user(s) of the results.

BACKGROUND INFORMATION

In recent years, the Internet has become more popular, allowing users to access information in ways never before possible. Information on the Internet is useful to a broad spectrum of individuals, including investors, financial analysts, entrepreneurs and academics. Changes reflected in a Web page, a large dip in a stock's value or a new merger amongst competing companies for instance, may be of critical importance to the users of the network.

However, keeping track of this constantly changing information is not easy. Logging on to the Internet every day to monitor changes can be time consuming. Thus, Web utilities have been developed, for instance FirstFloor's "Smart Bookmarks" (TM), which periodically monitor user defined Web sites and automatically notify the user when changes occur at a Web site (e.g., a change in the text), without the user having to log on.

Unfortunately, such monitoring utilities generally don't allow the user to customize the type of change or information being monitored to the particular interests of the user. In addition, while such utilities can
5 monitor changes within a particular Web site, they cannot monitor relationships amongst changes in several Web sites. Such changes, for example, the addition of a keyword reflecting a new product at several Web sites, may indicate the move to a new technology. Similarly,
10 newly created links amongst a group of Web sites may indicate an industry alliance is forming. Such industry wide (not company specific) information may be of greater importance to the user than changes reflected at a single Web site. Thus, a user may be forced to constantly monitor several Web sites in search of such new relationships, an even more time consuming and difficult task than monitoring a single Web site.

Another limitation of such utilities is that once a change is detected at a given Web site, the change is automatically communicated to the user, regardless of its importance. Thus, a great deal of information irrelevant to the user may nevertheless be communicated, again
wasting the user's time.

What is needed is a utility which (1) monitors the Internet for the occurrence of user defined conditions, including relationships amongst one or more Web sites,
(2) performs a search for information of importance to
30 the user upon detection of such conditions and (3) provides the user with the results of the search.

SUMMARY OF THE INVENTION

5 An object of the present invention is to overcome
the above-mentioned disadvantages of prior art Web
monitoring and search utilities. The present invention
(1) periodically monitors for user defined conditions,
including relationships amongst one or more network
sites, (2) probes for further information upon detection
of such conditions and (3) automatically reports results
10 to the user.

Thus, one advantage of the present invention is that
it automatically monitors for conditions and
relationships specified by a user, thus customizing the
automatic monitoring according to his or her interests.
15

Another advantage of the present invention is that
it allows automatic monitoring of changes and
relationships involving more than one Web site or Web
page. Such relationships (e.g., industry wide, as
20 opposed to company specific, trends) may be of particular
import to users.

Another advantage of the present invention is that
upon the detection of a user specified condition(s), the
system performs a user specified search for information
which may be of import to the user or which may explain
the occurrence of the user specified condition(s). In
other words, the present invention does not immediately
25 alert the user as soon as a minor change in the domain of
his interest takes place. The system continues to
explore to see if this change is accompanied by some
activities that are of interest to the user and returns a
more refined list of those changes. This probing feature
30

0001490-012698
000000-000000

provides at least two related benefits.

5 First, it acts as an information filter. Thus, it avoids the information overload that makes most simple monitoring mechanisms less than useful.

10 Second, the feature allows for "intelligent" information gathering from the Internet or other related networks. The present invention allows the user to define triggers that make an explicit connection between what is a symptomatic event and what needs to be further investigated. For example, the occurrence of an event (e.g., the deletion of a link to a site) may be significantly correlated with one phenomenon (e.g., addition of a link from a different site) but may be entirely meaningless in the context of other phenomenon (say, spurt in the number of visits to the site). Thus, the combination of a monitoring feature and probing feature allows an Internet user to find logical relationships between events, as well as learn their possible implications.

25 Another advantage of the present invention is that its monitoring, searching and notification functions may take place without a user log-on into the Internet.

30 Another advantage of the present invention is that a user may specify the monitoring conditions and searches in high-level language, similar to English. Thus, a user need not have specialized computer programming knowledge to benefit from the present invention.

Further objects, advantages and benefits of the present invention shall become apparent upon review of

the following detailed description and drawings below.

BRIEF DESCRIPTION OF DRAWINGS

5 Fig. 1 shows one possible apparatus implementing an embodiment of the present invention.

10 Fig. 2 is a flow chart depicting an overview of one possible embodiment of the method of the present invention.

15 Fig. 3 shows a general overview of the components of the monitor/probing rules which can be specified by a user and which is interpreted and executed by processing device in one possible embodiment of the present invention.

20 Fig. 4 depicts a possible syntax and component structure of an atomic condition which may be specified in the monitor probing rule depicted in Fig. 3.

25 Fig. 5 depicts a possible syntax and component structure of the literals shown in Fig. 4.

30 Fig. 6 depicts in detail a possible syntax and component structure of the predicates shown in Fig 5.

35 Fig. 7 is a flow chart showing steps which may be performed by an agent program and by a processing device at a monitoring site in accordance with instructions specified in the monitoring/probing rule depicted in Fig. 3.

Detailed Description of the Invention

Fig. 1 shows apparatus implementing one embodiment of the present invention. Client station(s) 1 are coupled via communications or other lines to monitoring site 2 which is coupled to (possibly) remote monitored sites via a network 11 (e.g., the Internet).

The client station(s) 1 comprises a computer processing device 3, storage device 4, memory device 12, display device 7, user input device 10 and user interface element 13. The computer processing device 3 can be, for example, a single microprocessor chip (such as an Intel Pentium chip), printed circuit board, several boards or other processing device. The storage device 4 can be, for example, an internal or external disk, Tape Cartridge or CD-ROM. The faster access and greater storage capacity the storage device 4 provides, the more preferable the implementation of the present invention. The memory device 12 can be implemented with, for example, a collection of RAM chips. The display device 7 can be implemented, for example, with a monitor, whether analog or digital. The user input device 10 can be implemented, for example, with a keyboard, mouse or scanner. The client station(s) 1 also include a user interface element 13, which can be implemented as a separate "software" (i.e., a program, process) whose instructions are executed by the computer processing device 3. However, there is no reason the computer processing device 3 (e.g., microprocessor chip) could not include the user interface element 13 process itself.

The monitoring site 2 comprises a computer

processing device 3, storage device 4, memory device 12, rule interpretation element 6, database allocation and search element 8 and agent creation/modification element 9. The computer processing device 3 can be implemented with, for example a single microprocessor chip (such as an Intel Pentium chip), printed circuit board, several boards or other device. The storage device 4 can be implemented with, for example, an internal hard disk, Tape Cartridge or CD-ROM. The faster access and greater storage capacity the storage device 4 provides, the more preferable the embodiment of the present invention. The memory device 12 can be implemented with, for example, a collection of RAM chips. The rule interpretation element 6, database allocation and search element 8 and agent creation/modification element 9 can be implemented as a separate "software" (i.e., programs, processes) whose instructions are executed by the computer processing device 3. Again however, there is no reason the computer processing device 3 (e.g., microprocessor chip) could not be used to implement the rule interpretation element 6, database allocation and search element 8 and agent creation/modification element 9 processes itself.

If the client station(s) 1 are connected to the monitoring site 2 via a telephone system or other system/lines not carrying digital pulses, the client station(s) 1 and monitoring site 2 may both also include a communications converter 5. The communications converter 5 can be implemented with any communications converter device, such as, for example, a modem. The communications converter 5 converts digital pulses into the frequency/signals carried by the line and also converts the frequency/signals back into digital pulses,

allowing digital communication.

As depicted in Fig. 1, the monitoring site 2 is connected to one or more client station(s) 1, and its resources may be shared by one or more clients in most common implementations of the present invention. Thus, for such implementations, the monitoring site 2 may be referred to as a "server" with "clients" (i.e., client stations 1), wherein the "server" is coupled to a (usually remote) network.

However, despite the embodiment described above, it is worthy of note that client station(s) 1 may each contain the elements included in the monitoring site 2. Therefore, a monitoring site 2 is unnecessary to implement the present invention.

Fig. 2 is a flow chart depicting an overview of one possible embodiment of the method of the present invention.

In Step 100 depicted in Fig. 2, a user at client station 1 specifies via user input device 10 one or more monitor/probing rule(s) using the facilities of the user interface element 13 which is displayed on the display device 7. Put simply, the monitor/probing rule specifies what information the user wants monitored, what information the user wants retrieved and under what circumstances. The user interface element 13 can be a set of choices in a displayed menu, a displayed form or other means for facilitating (preferably) easy, intuitive input.

In Step 110, the monitor/probing rule(s) data is sent from client station 1 to monitoring site 2. Again, the communication converter 5 may be necessary to effect this communication.

5

In Step 120, the rule interpretation element 6 running off the computer processing device 3 at the monitoring site 2 converts the monitor/probing rule data into a set of instructions the computer processing device 3 is already designed to execute.

10

The computer processing device 3 on the monitoring site 2 executes the interpreted instructions of the monitor/probing rule, which are in memory device 12 for this purpose. This execution results in two operations.

First, in Step 130, the database allocation and search element 8 running off the computer processing device 3 at the monitoring site 2 allocates space on the storage device 4. Conceptually, in one possible embodiment of the present invention, this can be thought of as creating space for a table(s) of records, with one or more fields per record. Initially, the conceptual table(s) are empty, but records are added to the table(s) later on as information is retrieved for monitoring purposes.

25

Second, in Step 140, the agent creation/modification element 9 also running off the computer processing device 3 at the monitoring site 2 alters a "template" agent program already existing in the storage device. In an alternative embodiment, the computer processing device 3

30

may create a new agent if one already doesn't exist.

Agents are well known in the art. Briefly, "agents" are programs which can run off the computer processing device 3 of a computer, be sent over communications lines and can then be run off the computer on the receiving end. They are therefore particularly suited for search and retrieval of information over the Internet.

In the above described embodiment of the present invention, once created/modified, the "agent" program periodically executes retrieval and monitoring operations. Also in a possible embodiment of the present invention, once user-specified monitored conditions are satisfied, the database allocation and search element 8 running off computer processing device 3 of the monitoring site 2 will execute a user-specified exploring request (e.g., search or retrieval operation) in storage device 4 through files therein (i.e., through conceptual table(s) of records retrieved by agent operations from Internet 11). Results of search can be sent from monitoring site 2 to client station(s) 1, and then can be displayed on display device 7 for user's benefit.

The method and apparatus of the embodiment heretofore described will become more apparent from the more detailed description below.

Note that the monitor/probing rules which may be specified by a user (Fig. 2, Step 100) are preferably in a "high-level" language (i.e., like English, or otherwise such that the user can write instructions without specialized programming language). In order to

accomplish this, the embodiment should preferably include means to convert the high level language in which the monitor probing rule is specified, into a set of instructions which may be executed by processing device of a computer.

For example, in the presently described embodiment, the monitoring/probing rule data is sent to monitoring site 2 (Fig. 2, Step 110) which includes rule interpretation element 6 (i.e., a language processor in the form of "software"). Rule interpretation element 6 perform the rule interpretation function depicted in Step 120, Fig. 2. Such a translator must necessarily encompass rules governing the structure of the higher level language being translated. A possible set of such rules is more fully described as follows.

Fig. 3 shows a general overview of the components of the monitor/probing rules which may be specified by a user (Step 100, Fig. 2) and which is interpreted and executed by a processing device in one possible embodiment of the present invention. User via user input device 10 specifies a statement of the form WHEN - IF - THEN, as shown in Fig. 3.

The WHEN and IF clause form the monitoring part of the rule trigger, and the THEN clause forms the probing part of the rule. The monitoring part of the trigger will effect a "watching" for certain events to occur that satisfy certain conditions. Once the present invention detects appropriate events, it does not necessarily inform the user about this. Instead, the present invention tries to explore the situation in which these events occur by executing certain probing actions

specified in the probing part of the trigger i.e., the THEN clause. The structure of a rule is similar to the structure of triggers used in active databases; however, the nature of the events, conditions and especially probing actions is substantially different, as will be explained below.

In the case of the WHEN portion of the statement, user specifies an "event" parameter (Step 200). An event is a change in the state of one or more of the objects in (e.g., Web-sites on the Internet 11) that are being monitored. For instance, an appearance or disappearance of one or more keywords, the appearance of a new link between two sites, a change of physical attributes of a Web page or the like are examples of events.

In the case of IF portion of the statement, the user specifies a "condition" parameter (Step 210). Conditions are constraints that the user applies to events and act as filters that refine the space of the events to result in a smaller subset, that are of interest to the user.

In the case of the THEN portion of the statement, the user specifies a "probing action" parameter (Step 220). In general, the probing action investigates what is going on at the site or a collection of sites once the events in the WHEN clause occurred and satisfied the conditions of the IF clause.

The following examples illustrate the general structure of the WHEN - IF - THEN monitor/probing rule(s), translated in plain English.

Example 1: WHEN the Keyword "cable modem" appears on

a site IF the Keyword "push technology" is not included in the text of the site THEN find out how many sites from the predefined list of sites the word "Cable modem" appeared and of these how many featured the Keyword "Push technology".

Example 2: IF several Web-sites are linked into a "star" configuration AND the Keyword "cable modem" is found on all the nodes in the configuration THEN find all such "star" configurations.

Example 3: IF a Keyword consisting of "Push technology", "Web TV", "Channels" appears in all Nodes (members) of a user defined configuration THEN find all such configurations where such Keywords appeared and return the list of the corresponding Uniformed Resource Locators (URL's).

Example 4: IF the number of visits to a site increase at a rate greater than **R** AND the sites contain the Keywords set {Cable modems, Web TV} THEN find all sites that contain the same Keywords and whose visit rate was greater than **R**; return the set of corresponding URL's.

Example 5: WHEN a Keyword appears at a site AND links to that site from other sites disappear THEN find all the sites from which links to this site disappeared, find if these sites have any Keywords in common, find all the sites which are now linked to this site, find if these sites have any Keywords in common and return the four lists.

Notice from the above examples also that although

events in the monitoring part of the rule are useful, it is not necessary to use them in general for monitoring Web activities. The reason for that is that they can be simulated through the conditions of the IF clause alone (the events describing a set of circumstances which are a subset of those described by the conditions and events). Also, a monitoring/probing rule may contain only the WHEN and THEN clauses (i.e., no IF clause).

For simplicity sake, in the embodiment described more fully below events in the WHEN clause (Fig. 3, Step 200) will be specified by indicating certain conditions in the IF clause (Fig. 3, Step 210) using the methods known to a person having skills in the art. Therefore, we will focus only on the structure of conditions and probing actions in the rest of this section. Again, the present invention is by no means limited to the embodiment described below.

Note that an event may be atomic or composite. The composite event is the conjunction of two or more atomic events. One of the most popular methods to define composite events is through conjunction of atomic events, *i.e.*, as E1 and E2 and...En, where E1 and E2 and ...En are atomic events, such as "a link between two sites appears," "a Keyword appears within a site," or "a page or link was deleted." However, composite events can be defined in more general terms as well (*e.g.*, as a conjunctive or disjunctive form (CNF or DNF) of its atomic events). Atomic events may belong to one of the following classes (but atomic events are not limited to just these classes): 1) a link appears/disappears; 2) a Keyword (or a group of Keywords) appears on a page; 3) a page's text is modified; 4) change in the physical

attributes of a page; 5) a visitor visits a page

One way to specify conditions in the IF-clause of a monitoring/probing rule would be through a conjunction of atomic conditions, i.e. as P_1 and P_2 and P_n , where P_i is an atomic condition. Consistent with this method, in general, the IF clause can be defined as the conjunctive normal form or a disjunctive normal form of atomic conditions. Conjunctive and disjunctive normal forms are well known in the art. An example of an IF-clause is: IF Site S_1 is Linked to Site S_2 AND Keyword K_i appears in S_2 AND Keyword K_j exists in S_1 (where "Site S_1 is Linked to Site S_2 ," "keyword K_i appears in S_2 " and "Keyword K_j exists in S_1 " are atomic conditions).

In order to define atomic conditions, it is also necessary to explore the modeling primitives (or the basic entities) used in our monitored environment. In one possible embodiment of the present invention, the information to be monitored, searched and retrieved is on the Internet, or more specifically the World Wide Web. In this environment, the modeling primitives (basic entities) might be:

1. **Web Pages:** A web page is a collection text, pictures, video and audio objects (and other media) that are on the Web and share the same URL. Two different Web pages will not have the same URL. A Web page may have physical attributes such as size, date of last change, title etc. Web pages may also be referred to as 'Pages' below.
2. **Web Sites:** A Web Site is a collection of one or more Web pages. Pages within a site are linked by links.

A Web Site also has a URL which uniquely identifies the site. Web sites may also be referred to as Sites.

3. **Key Words:** These are text objects (words, phrases etc.) found in a web page and therefore, by extension found in a Web site.
4. **Links:** Links are embedded in Web pages and allow a visitor to move from one location to another. A Link may take a user from one page to another or may simply take her to another location on the same page.
5. **Visit:** Refers to the act of a visitor (anybody who reaches a Web Site) visiting a site.

In addition to these basic entities, composite conditions can be derived from the basic entities using a data definition language (for example, if the basic entities are defined in terms of a relational data model as tables, the composite entities are defined as relational views, i.e., as SQL expressions). Examples of these composite entities are:

1. **Configuration:** A collection of Web Sites that are linked to each other. A Configuration can have one (trivial case) or more Web Sites. For example, a set of sites linked in a circular fashion forms a "ring" configuration; or a configuration consisting of a central site that is linked to all other sites in the configuration and other sites being linked only to the central site, forms a "star" configuration.

2. **Reach relation:** It specifies which sites can be reached from a given site. This relation can be defined as a transitive closure of the LINKS relation.

5

In practicing the present invention, basic and derived entities are modeled as data objects in the corresponding data model. For example, if the data model is relational, then these entities are modeled as relations (predicates). If the data model is object-oriented, then they are modeled as data objects.

10

00013490012690
0692T01064ET060

In the embodiment described below the relational data model is used in building the rule interpretation means 6 and database allocation and search means 8 and the basic entities are modeled as predicates (using relational terminology, they can sometimes be called "relations" or "tables"). For example, LINKS is a predicate of the form LINKS(site1,site 2), where site1 is the site from which the link originates and points to site 2. However, the present invention is not limited to the relational data model and is applicable to other data models as well.

20

Moreover, some of the predicates are temporal because they change over time. For example, the predicate LINKS may change over time (because links can

25

be added and removed over time).

Once the basic entities of the data model are defined (e.g., as predicates/relations), the atomic conditions in the IF clause of the monitor and its structure can be defined as illustrated in Fig. 4. Fig. 4 depicts a possible syntax and component structure of atomic conditions.

As depicted in Fig. 4, an atomic condition can either be a literal, or a past unary temporal operator followed by a literal or a literal followed a binary past temporal operator followed by a literal. Some examples of unary past temporal operators include "Always_in_the_past," "Sometimes_in_the_past," "Sometimes_within_the_past(T time units)" and "Always_within_the_past(T time units)." An example of a binary temporal operator is "Since."

As depicted in Fig 5, a literal is either a predicate or a negation operator (e.g., "NOT") followed by a predicate.

Fig. 6 depicts in detail a possible syntax and component structure of the predicates shown in Fig 5. Again, Fig 6. depicts one possible syntax and component

structure of such predicates, and the present invention is in no way limited thereto. As depicted in Fig. 6, a predicate may be constructed by one of five expressions followed one or more parameters in parentheses. Based on the basic entities of the Internet environment described above, the expression may be "WebPages," "WebSites," "KeyWords," "Links" and "Visit."

Based on the rules described above, some examples of atomic conditions which may appear in the IF portion of the monitoring/probing rules is as follows:

1. Key_Words(site,keyword,count), a predicate meaning that the user has specified that the cite keyword "keyword" has count "count" at site "site".

2. "LINKS(site1,site 2) Sometimes_within_the_past(2weeks)" -- meaning that a link existed at some point between sites site1 and site 2 within the past 2 weeks. Note that LINKS is a temporal predicate because it implicitly changes over time.

3. "Key_Words(XYZ, "cable - modem", count) Since LINKS(XYZ, ABC)" -- meaning that the keyword "cable modem" appeared at cite XYZ since the link was established between sites XYZ and ABC.

Turning now to the THEN clause portion of the monitoring/probing rule, the THEN clause consists of a sequence of probing operations. In one possible embodiment of the present invention, the structure of the THEN clause can be defined as a query in Structured Query Language ("SQL"). SQL is well known in the art and is suited to search and retrieval in a relational database environment, such as the one described above. Thus, the rule interpretation means 6 may include SQL interpretation means, well known in the art. However, the THEN clause can include numerous other constructs some of which are discussed in greater detail later.

Examples of probing actions expressed as SQL queries follow. Assume that in the monitoring part of a trigger it was detected that a certain configuration among a group of sites was broken when a certain keyword appeared in one or more of these sites. Then in the probing part we may want to find the list of all sites whose links to those sites where the keyword appeared were deleted after the appearance of the keyword. Such probing actions can be expressed with SQL queries, three examples of which follow.

The first query extracts the list of all sites that point to one particular site (e.g., www.target.com) in the previous monitoring period, before the keyword appeared:

```
5      SELECT [Web Links Data]. [Origin of Link], [Web Links
Data]. [Target Site],
        [Web Links Data]. [Monitoring Period], [Web Links
Data]. Status
10     FROM [Web Links Data]

WHERE ((([Web Links Data].[Target
Site])="www.target.com") AND (([Web
Links Data]. [Monitoring Period])="T - V) AND (([Web
Links Data]. Status) =Yes)).
```

The second query extracts all those sites that do not have a link with that site in the present monitoring period and stores it as a temporary table called ("TEMP: Sites with no current link"):

```
25     SELECT [Web Links Data]. [Origin of Link], [Web Links
Data]. [Target Site],
        [Web Links Data]. [Monitoring Period], [Web Links
Data]. Status

FROM [Web Links Data]

30     WHERE ((([Web Links Data].[Target
Site])="www.target.com") AND (([Web
Links Data]. [Monitoring Period])="T") AND (([Web
Links Data].Status)=No)).
```

The third query joins the two temporary tables in an SQL Query and extracts the list of those sites that have deleted a link in the current monitoring period.

```
SELECT [TEMP: Sites with no current link]. [Origin of
Link]
```

FROM [TEMP: Sites with no current link] INNER JOIN
 [TEMP:sites with earlier
 link] ON [TEMP: Sites with no current link].[Origin
 of Link] = [TEMP:sites with earlier link]. [Origin
 of Link]
 GROUP BY [TEMP: Sites with no current link]. [Origin of
 Link]
 ORDER BY [TEND: Sites with no current link]. [Origin of
 Link].

Thus, the WHEN-IF-THEN, or as described above, IF-
 THEN rule, can be described by constructs such as those
 described above as well as, for example, SQL
 interpretation means and thus translated by rule
 interpretation element 6 into instructions executed at
 monitoring site 2 by computer processing device 3.

The computer processing device 3 executes the
 interpreted instructions of the monitor/probing rule,
 which are in memory device 12 for this purpose. In the
 embodiment described herein, this execution results in
 two operations.

First, turning again to Fig. 2., in Step 130, the
 database allocation and search element 8 running off the
 computer processing device 3 at the monitoring site 2
 allocates space on the storage device 4. Conceptually,
 this can be thought of as creating space for a table(s)
 of records, with one or more fields per record.
 Initially, the conceptual table(s) are empty, but records

are added to the table(s) later on as information is retrieved for monitoring purposes.

5 For example, assuming that the IF portion of a monitor/probing clause consists entirely of an atomic condition in the form Key_Words(site, keyword, count), a file is created on the monitoring site 2 which can be conceptualized as a "Key_Words" table with "site," "keyword" and "count" fields (or columns). Each time a new keyword entry at a site is retrieved from the Internet, the entry can be conceptualized as another row of the table which includes the appropriate count of the number of times this keyword appears at that site.

10
15
20 Alternatively, assuming that the IF portion of a monitor/probing clause consists entirely of an atomic condition "Links(site1, site2) Sometimes_within_past(2weeks)," a file may be created on the monitoring site 2 that is represented as a temporal relation using methods well known in the art. Alternatively, such temporal atomic conditions may be represented on the monitoring site 2 using the implementation methods described later.

25 Next, in Step 140 of Fig. 2, the agent

creation/modification element 9 also running off the
computer processing device 3 at the monitoring site 2
alters a "template" agent program already existing in the
storage device 4. In an alternative embodiment, the
5 computer processing device 3 may create a new agent if
one already doesn't exist.

A more detailed description of Steps 130 and 140 is
as follows: without loss of generality, in the
following description, assume that triggers have only the
IF clause and the THEN clause (no WHEN clause).

First, determine the set of (temporal) predicates
for each atomic condition in the IF clause (either
temporal predicates or the auxiliary predicates as
described later). Allocate data storage for these tables
using the methods known in the art.

Second, given an IF-clause, determine for each
20 temporal or non-temporal predicate generated in Step 1
("real" predicate, not the auxiliary predicate) the set
of Web sites S from which information should be obtained
to evaluate that predicate. For example, if the IF
statement is

__ IF KeyWords(site,word,count) and site IN
{www.disney.com,www.oracle.com}

5 and predicate is KeyWords(site,word,count) then the set S
consists of {www.disney.com,www.oracle.com} for that
predicate. In this simple example, there is only one
predicate in a clause. In case there is more than one
predicate in the IF clause, the determination of the Set
of sites S can be done by someone with the ordinary skill
in the art.

Third, for each predicate and each site in set S
computed in Step 2, generate an agent that goes to that
site and periodically downloads the necessary data needed
to evaluate that predicate (note that the agent is
permanently allocated to the monitored site; once in a
while, it can be moved out by the site, but then it tries
to get back again).

20 The agent is developed as follows. An agent
template is written for each type of predicate by the
system developer at the time the monitoring system is
designed (this means that the Web monitoring system is
shipped to the user with a set of predefined agent
25 templates - that can be extended later by the user). For
example, for the LINKS predicate, the agent template is a

software_program that travels to a Website (specified as an argument for that template) and periodically downloads all the links to which this site points (note that the downloading period is another argument). When the agent
5 template is initiated as a part of the agent.

generation/sending procedure, the arguments of the template are initiated with real parameters. For example, in the case of the LINKS predicate, the agent template is instantiated with the parameters indicating the site to which the agent must travel and the downloading period. In the most general case, every monitoring period, the agent brings back to the monitoring site the entire predicate from that monitored site. For example, the "keywords" agent at the site "site", sends back to the monitoring site all the keywords and their counts from that site (i.e. that are parts of predicate KeyWords for that site).

However, certain optimization methods may minimize
20 the amount of data sent by the agent to the monitoring site. For example, if nothing changed at a site from the previous monitoring period, there is no need to transmit any data back to the monitoring site.

25 Once the data arrives to the monitoring from the monitored site, it is stored in the tables allocated in

Step 1 using the methods known to a person with an ordinary skill in the art.

5 In the above described embodiment of the present invention, once created/modified, the "agent" program periodically executes retrieval and monitoring operations (Fig. 2, Step 150) as follows.

Turning now to Fig. 7, it depicts a flowchart showing steps which may be performed by an agent program and by database search and allocation element 8 at monitoring site 2 in accordance with instructions specified in the probing part of the monitoring/probing rule depicted in Fig. 3.

99013490-0115998
20 In Step 300, at the first monitoring period, the Agent program is communicated from the monitoring site 2 to monitored site(s) via the network 11 (e.g., the Internet). Once at the monitored site(s), Agent program evaluates data, sends information back to monitoring site 2 based on user-defined criteria specified in the IF portion of the monitoring/probing rule and may remain at monitored sites until completion of its tasks (if the agent is terminated by the monitored site, another copy of the agent attempts to get back to the site. That copy

25

may come from either the monitoring site or elsewhere in the pipeline between the monitoring site and the monitored site).

5 For example, assuming that the IF portion of a monitor/probing clause consists entirely of an atomic condition in the form Links(site1, site2), the agent program retrieves the URL's corresponding to relevant first sites and second sites back to monitoring site 2. More specifically, an agent program may be sent to each of the relevant sites corresponding to the first argument of Links (i.e., Links(site1)). Each of the agents (sitting at site i) then sends back URL's of all the sites linked to the site at which the agent is sitting.

1593510"064ET005
13490"015698
20 The database search and allocation element 8 running off computer processing device 3 on the monitoring site 2 stores the URL data into a file on the storage device 4. This file can be conceptualized as a "Links" table. Each link can be defined by URL's of site1 and site 2. Thus, the table can be conceptualized as a table of columns "site1" and "site2," with each new Link entry forming a new row. In addition to this simple case, if the IF clause contains a temporal LINKs condition, the past history of links between sites would also be stored as

25

discussed in more detail later.

5 One main implementation issue relating to applications using the World Wide Web is that it is stateless in the sense that it does not "remember" its history. For example, there is no way to test whether the volume of visits at the site has been monotonically increasing within the past 3 months unless the client station 1 or monitoring site 2 maintains the information about the volume of visits over time (i.e., the Web site does not maintain this information itself). Therefore, it is the responsibility of the client station 1 or the monitoring site 2 to maintain the necessary information. Thus, two related implementation issues arise as a result of the agent's data retrieval from the Internet 11.

20 The first issue involves what historical information needs to be maintained and where for monitoring purposes.

25 Given a set of triggers, monitoring site 2 or client station 1 needs to maintain the (temporal) predicates that would allow processing device 3 to evaluate temporal expressions contained in these triggers . This means that

monitoring site 2 or client station 1 (hereinafter
monitoring site 2 alone will be referred to) must store
the current values of these predicates and, in certain
cases, their past histories. For example, assume that in
the IF clause user specified predicate "LINKS(site1,site
2) within_past(2months)" specifying that a link existed
between site1 and site 2 at some time within the past 2
months. To be able to monitor and evaluate such a
predicate, monitoring site 2 needs to maintain the
current state of predicate LINKS (at the present time) as
well as its past history within the last two months.
Since time is continuous, and processing device 3 can
evaluate the presence of links between site1 and site 2
only at some finite set of times, this means the user
needs to specify the sampling (monitoring) rate for the
evaluation of predicate LINKS. For example, the user may
specify to check the status of predicate LINKS every day,
or every 8 hours, or every hour. The Agent will then
retrieve information, and the information will be
checked, at the time points corresponding to the user-
specified sampling rate.

Notice that it can be assumed that no abnormal
changes happen to predicate LINKS between the sampling
points. For example, assume that agent samples predicate

LINKS every day. Then if the link between site 1 and site 2 existed yesterday and still exists today, this means that nobody deleted and inserted the same link within the last day (but it is of course possible to delete the link during that day; in this case, the present invention detects the change at the next sampling point).

Rather than storing the whole past history of temporal predicates, it is more efficient to store the temporal information in compressed format using the methods described below. To illustrate how this can be done, consider the example of the temporal predicate "always_past LINKS(site 1, site 2)," stating that there was always a link between site 1 and site 2 in the past. In that case, the monitoring site 2 may maintain two predicates LINKS(site 1, site 2) and Always_Past_Links(site 1, site 2) in order to monitor condition "Always_Past Links(site 1, site 2)", where Always_Past_Links is updated each monitoring period by removing from it the records that do not appear in predicate LINKS .

In general information to be maintained at the monitoring site 2 can be determined as follows. Take the set of all the temporal predicates appearing in the

00013490-015000

monitoring parts of all the triggers. If the predicate is non-temporal (no trigger refers to its past history), database search and allocation element 8 will store the current copy of that predicate on the storage device 4 of the monitoring site 2 (the copy at the latest sampling period). If the predicate is temporal (some triggers refer to its past history), database search and allocation element 8 stores the current copy of that predicate and the following additional information. For each temporal operator (e.g. Always_past, Sometimes_past, etc.) database search and allocation element 8 store an auxiliary copy of the predicate pertaining to this operator. For example, for the temporal expression Always_past LINKS(site1, site 2) appearing in one of the triggers, database search and allocation element 8 store the current copy of LINKS(site1, site 2) and an auxiliary predicate Always_Past_Links(site1, site 2). Always_Past_Links(site1, site 2) is true if and only if there was always a link from site1 to site 2 in the past. Similarly, we store an auxiliary predicate Sometimes_Past_Links for the temporal expression "Sometimes_Past Links." In general, storage device 4 at monitoring site 2 maintains one auxiliary operator for every (predicate, temporal operator) pair appearing in one of the monitoring triggers (e.g., IF clause) created

by the user. These auxiliary predicates can be maintained using the methods known to the person having an ordinary skill in the art (e.g. as described in Chomicki, J., "History-Less Checking of Dynamic Integrity Constraints," 8th IEEE Conference on data Engineering, Phoenix, Arizona, 1992, p. 25-26).

For example, predicate (table) Always_Past_Links is updated by database search and allocation element 8 every monitoring period by removing records (site 1, site 2) from it that do not appear in the current copy of Links(site 1, site 2). Note that these auxiliary predicates are *non-temporal* (they only simulate temporal predicates but do not have any temporal component as part of their structure) and can be stored using the conventional methods known to someone with an ordinary skill in the art.

The retrieval of information from the Internet also raises a second, related issue. One of the steps in predicate maintenance noted above is the step of agent obtaining the current copy of the predicate. This means the agent must visit the monitored site(s) (e.g., on the Internet 11) and evaluating the present conditions of the predicate(s) at that site. One of the central issues in

00013490-01693
359310-064510

this process is what information should the agent check and, even more importantly, what information it should bring back from the monitored site(s) 11 to the monitoring site 2. This is a critical issue because, we want to monitor many sites and need to minimize the data to be transferred back to the monitoring site in order not to clog the communication lines.

As was demonstrated above, for the monitoring purposes, it is necessary for the agent program to bring back at each monitoring period only the current states of the predicates that are maintained on the storage means 4 of the monitoring site 2, such as Links, Web_Sites, Key-Words, etc.

An optimization issue arises when less data than is in the current copy of the predicate being monitored can possibly be brought back to the monitoring site 2. For example, assume that the temporal predicate being monitored in the IF-clause of a trigger is "Always_Past LINKS(site 1, site 2)". As was explained above, this predicate can be monitored through the database search and allocation means 8 creating and storing an auxiliary predicate Always_Past_Links(site 1, site 2). Note that this predicate is monotonically decreasing over time.

Also note that rather than retrieving the relation LINKS in its entirety at each monitoring period, the agent program may retrieve only the records from the intersection of Always_Past_Links and Links.

5

In general, if auxiliary predicates are monotonically decreasing (such as Always_past_Links), then the agent retrieves auxiliary predicates rather than the current predicates (e.g. LINKS). To the contrary, if auxiliary predicates do not decrease over time (such as Sometimes_Past_Links), the current predicates are retrieved by agent from the Web sites on the network 11 being monitored.

Data transmission can be further optimized when no significant changes occur at the monitored sites. In that case, data need not be transmitted to the monitoring site 2. For example, the agent may have stored a previous copy of the predicate either at monitored or, in the event this is impossible, at other intermediate sites in the network 11. To determine whether a predicate has changed since the last monitoring period, the agent may then compare the previous copy with a current copy of the predicate using techniques known in the art. If the copies differ then, and only then, is

the predicate transmitted to the monitoring site 2.
Alternatively, agent may determine the differences
between current and previous copies and transmit only
those differences.

5

The optimization and implementation issues noted
above are effected by agent creation/modification element
9 at monitoring site 2.

Turning now to Fig. 7, Step 302, in one embodiment
of the present invention, once the Agent program
retrieves user-specified data to the monitoring site 2
and database search and allocation element 8 stores the
information, processing device 3 checks the stored
information to determine if it satisfies the user-
specified condition (e.g., in IF clause) (using methods
well known in the art).

In one embodiment of the present invention, when the
user-specified condition is satisfied, then in Step 303
(Fig. 7), database search and allocation element 8
running off processing element at 3 monitoring site
performs SQL search on files (tables) of data on storage
device 4 already retrieved from Internet 11, as described
above, or performs other probing actions described later.

5 In Step 304, the results of the SQL search or other
probing actions are sent from monitoring site 2 to client
station(s) 1 via e-mail (or other notification means)
where they may be displayed on display means 7 for the
user's benefit. E-mail is commonly used and its
implementation is well-known in the art. Of course such
notification becomes even simpler when the present
invention is practiced without monitoring site 2, as
noted above.

10
15
20 Notice that in an alternative embodiment, the
probing operation may search and return information from
the Internet 11 instead of monitoring site 2 as just
mentioned. Where information is returned from the
Internet 11, one of ordinary skill may use any of the
existing known Web-related information retrieving
methods.

25 In addition, in other alternative embodiments, other
types of probing or other operations may also be
performed as a part of the THEN-clause upon satisfaction
of user-defined conditions of the IF-clause

As a first example, monitoring site 2 may send back to client station 1 information corresponding to user-defined condition detected by processing means 3 at monitoring site 2 and client station 1 simultaneously issues alarm. For example, the IF-THEN clause (in English):

IF the keyword "Cable Modem" appears at a site but the keyword "Push Technology" does not appear at that site THEN issue an alarm and return that information to the user,

returns the names of all the sites where the keyword "Cable Modem" appears but not the keyword "Push Technology."

As a second example, the present invention may execute a program written in a general-purpose programming language (e.g. C) that examines certain conditions. For example, such program may visit different Web sites in some sequence and return the URL's of those sites that satisfy certain properties (e.g. have increasingly higher visitation traffic).

As a third example, a Data Mining Query may be

executed. Data Mining queries, also known as pattern templates, are constructs for the specification of types of patterns to be discovered. Therefore, probing actions can further explore what is "going on" on one or several network sites when the IF condition(s) of a rule are satisfied by letting the Data Mining Query discover network-related patterns. Several researchers considered data mining queries (pattern templates) in the context of knowledge discovery. Imielsinski, T., Virmani, A. and Abdulghani, A., "DataMine: Application Programming Interface and Query Language for Database Mining, "Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, August, 1996; Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H. and Verkamo, A.I., "Finding Interesting Results from Large Sets of Discovered Association Rules, "Proceedings of the Third International Conference on Information and Knowledge Management," December, 1994; Han, J., Fu, Y., Wang, W., Koperski, K. and Zaiane, O., *DMQL - A Data Mining Query Language for Relational Databases*, "Proceedings of the (SIGMOD) Workshop on Research Issues on data Mining and Knowledge Discovery," June, 1996; Shen, W., Ong, K., Mitbender, B. and Zaniolo, C., *Metaqueries for Data Mining*, "Advances In Knowledge Discovery and Data Mining," chap. 15, AAAI Press, 1996.

Imielsinki, T., Virmani, A. and Abdulghani, A., "*DataMine: Application Programming Interface and Query Language for Database Mining*, "Proceedings of the Second Interbational Conference on Knowledge Discovery and Data Mining, August, 1996 introduce the M-SQL query language as an extension of SQL by including a small set of primitive data mining operations. M-SQL queries operate on the underlying database and on a set of previously discovered rules and return a set of rules discovered in the data that satisfy the conditions of the query.

For example, the query "Find all rules in Table T involving attributes Disease, Age, and ClaimAmt, which have a confidence of at least 50\%" can be expressed in M-SQL (Imielsinki, T., Virmani, A. and Abdulghani, A., "*DataMine: Application Programming Interface and Query Language for Database Mining*, "Proceedings of the Second Interbational Conference on Knowledge Discovery and Data Mining, August, 1996) as:

```
SELECT      *
FROM        Mine(T) R
WHERE       R.Body < {(Disease=*), (Age=*),
                    (ClaimAmt=*)}
and {} < R.Body and R.Consequent IN
```


-- {(Disease=*), (Age=*), (ClaimAmt=*)} and
R.Confidence > 0.5

5 The Mine operator returns all the association rules
(Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and
Verkamo, A.I., "Fast Discovery of Association Rules,
"Advances in Knowledge Discovery and Data Mining," Chap.
12, AAAI Press, 1996; Imielsinski, T., Virmani, A. and
Abdulghani, A., "DataMine: Application Programming
Interface and Query Language for Database Mining,
"Proceedings of the Second International Conference on
Knowledge Discovery and Data Mining, August, 1996) that
are true in T, and the WHERE-clause imposes conditions on
them, such as that the body of a rule should not be empty
and should contain attributes Disease, Age, and ClaimAmt.
Also, the consequent of the rule must contain one of
these attributes and the confidence should be at least
50\%.

20 M-SQL queries are evaluated by launching discovery
methods, as described in *DataMine: Application
Programming Interface and Query Language for Database
Mining*, "Proceedings of the Second International
Conference on Knowledge Discovery and Data Mining,
August, 1996.

Han et al. (in Han, J., Fu, Y., Wang, W., Koperski, K. and Zaiane, O., *DMQL - A Data Mining Query Language for Relational Databases*, "Proceedings of the (SIGMOD) Workshop on Research Issues on data Mining and Knowledge Discovery," June, 1996) designed the data mining language DMQL. A query in DMQL consists of the specification of four data mining primitives:

- (1) the set of data relevant to the data mining process;
- (2) the kind of knowledge to be discovered, i.e. a characteristic, discriminant, classification, or an association rule;
- (3) the background knowledge consisting of a set of concept hierarchies or generalization operators which provide corresponding higher level concepts;
- (4) the interestingness of the knowledge to be discovered, which is specified with a set of different mining thresholds.

For example, the query "Classify students according to their GPA's and find their classification rules for those majoring in computer science and born in Canada, with the attributes birth_place and address in consideration" can be expressed (Han, J., Fu, Y., Wang, W., Koperski, K. and Zaiane, O., *DMQL - A Data Mining*

Query Language for Relational Databases, "Proceedings of the (SIGMOD) Workshop on Research Issues on data Mining and Knowledge Discovery," June, 1996) as:

5 Find classification rules for CS_Students
According to GPA
Related to birth_place, address
From Student
Where major = 'CS' AND birth_place = 'Canada'

10 It was Shen et al. (in Shen, W., Ong, K., Mitbender,
15 B. and Zaniolo, C., *Metaqueries for Data Mining*,
"Advances In Knowledge Discovery and Data Mining," chap.
15, AAAI Press, 1996) that define metaqueries, which can
be viewed as a two-part specification. The left-hand
side (LHS) of a metaquery specifies a constraint on how
data should be prepared, and the right-hand side (RHS)
specifies an action to be applied on the prepared data.
For example, consider metaquery:

20 $P(X,Y) \text{ AND } Q(Y,Z) \rightarrow R(X,Z)$

25 This metaquery is a template that instantiates various
transitivity patterns depending on the predicates to be
substituted for templates P, Q, and R and the attributes

and values substituted for variables X , Y , and Z .

For instance, consider the transitivity pattern:

5 $p(X,Y) \text{ AND } q(Y,Z) \rightarrow r(X,Z)$ [with probability Pr]. This pattern satisfies the earlier specified metaquery, where predicates p , q , and r are specific database relations, and Pr is the ratio of the (X,Z) pairs satisfying the LHS and the RHS of the rule and those satisfying only the LHS. Then, the action in the RHS of the metaquery lies in computing the value of Pr .

The notion of a pattern template was also introduced in Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H. and Verkamo, A.I., "Finding Interesting Results from Large Sets of Discovered Association Rules," Proceedings of the Third International Conference on Information and Knowledge Management," December, 1994 for the purpose of identifying interesting patterns. A pattern template in Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H. and Verkamo, A.I., "Finding Interesting Results from Large Sets of Discovered Association Rules," Proceedings of the Third International Conference on Information and Knowledge Management," December, 1994 is a rule $A_1 \text{ AND } \dots \text{ AND } A_k \rightarrow A_{k+1}$

25 where each A_i is either an attribute name, a class name

or an expression C^+ or C^* corresponding, respectively, to one or more and zero or more instances of the class C . Such a template defines a class of rules that are instances of the template pattern.

5

For example, consider a pattern template that says that students can take an advanced elective in some department if they have taken a core course offered by that department and one or several electives:

CORE(Dept, Student) AND ELECTIVE(Dept, Student)+ -->
ADV_ELECTIVE(Dept, Student)

where CORE, ELECTIVE, and ADV_ELECTIVE are classes of certain courses offered at that university. Given this template and assuming that a and b are confidence and support thresholds (Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H. and Verkamo, A.I., "Finding Interesting Results from Large Sets of Discovered Association Rules, "Proceedings of the Third International Conference on Information and Knowledge Management," December, 1994), the rule

Intro_CS(CS, Student) AND Programming_Lang(CS, Student) AND
S

Operating_Systems(CS,Student-->

Special_Topics_in_OSes(CS,Student) [a, b]

matches this template.

5

The above references to user notification and program execution operations upon satisfaction of a condition can be implemented by a person having ordinary skills in the art. In addition, whenever information is to be returned from the monitoring site 2 to the user, one of ordinary skill in the art can use any of the known database information retrieval methods, such as the utilization of any of the existing database query languages.

It is readily apparent that numerous other modification, combinations of and minor additions to the above disclosure may be made for the benefit of network users while remaining within the scope of the present invention. For example, a useful way of implementing the invention is for the monitoring site 2 to act as server to more than one client station 1, and to maintain a separate account file for each such client station 1. The account files could include the monitoring conditions and probing instructions associated with their respective

client station. Accordingly, more than one client at more than one location could simultaneously benefit from the present invention, each according to their own needs, in the manner described above.

5

09013490-012698